



Free-Energy Based All-Atom Protein Folding Using Worldwide Distributed Computational Resources

T. Strunk, S. M. Gopal, K. Klenin, W. Wenzel

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 381-384, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Free-Energy Based All-Atom Protein Folding Using Worldwide Distributed Computational Resources

**Timo Strunk^{1,2}, Srinivasa Murthy Gopal¹,
Konstantin Klenin¹, and Wolfgang Wenzel¹**

¹ Institute of Nanotechnology, Research Centre Karlsruhe,
76344 Eggenstein-Leopoldshafen, Germany
E-mail: {Timo.Strunk, wenzel}@int.fzk.de

² Fachbereich Physik, Technische Universität Dortmund,
Otto-Hahn-Str. 4, 44227 Dortmund, Germany

We have implemented massively parallel stochastic optimization methods for all-atom de-novo protein folding using our free-energy forcefield PFF02^{1,2}, which is based on Anfinsen's thermodynamic hypothesis³. We have implemented this approach (POEM) using a world-wide volunteer computational grid to predictively and reproducibly fold the HIV accessory protein 1F4I from completely unfolded conformations.

1 Motivation

Protein folding and structure prediction at the all-atom level remain important computational challenges. To achieve this goal in the long term it is important to develop methods that are capable to predictively fold proteins and peptides from unbiased unstructured conformations to the native ensemble. Direct simulation studies have demonstrated the folding of several small peptides and mini-proteins from completely extended conformations, but remain limited in the system size by the large computational effort required.

One great hope towards reproducible all-atom folding is the development of algorithms that can exploit emerging massively parallel computational architectures. We have recently developed an evolutionary algorithm, which generalized the basin hopping or Monte-Carlo with minimization^{4,5}, method to many concurrent simulations. This approach was ported to the massively parallel BOINC architecture on POEM@HOME (<http://boinc.fzk.de>) and verified by folding several proteins reproducibly.

2 Method

We have parameterized an all-atom free-energy forcefield for proteins (PFF01/02)^{1,2}, which is based on the fundamental biophysical interactions that govern the folding process. We could show that near-native conformations of several proteins correspond to the global optimum of this forcefield. We have also developed, or specifically adapted, efficient stochastic optimization methods (stochastic tunnelling, basin hopping, evolutionary algorithms) to simulate the protein folding process. Forcefield and simulation methods are implemented in the POEM (Protein Optimization with free-Energy Methods) program package.

2.1 Optimization Strategy

We have generalized this method to a population of fixed size which is iteratively improved by an arbitrary number of concurrent dynamical processes^{6,7}. The whole population is guided towards the optimum of the free energy surface with a simple evolutionary strategy in which members of the population are drawn and then subjected to a single simulated annealing basin hopping cycle. At the end of each cycle the resulting conformation either replaces a member of the active population or is discarded. The decision tree for this process is illustrated in figure 1.

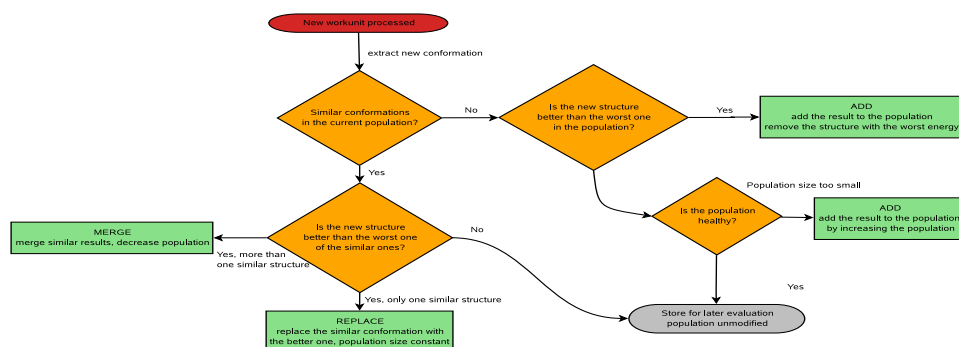


Figure 1. Evolutionary optimization strategy.

2.2 POEM@HOME

In this investigation, we deployed the BOINC server POEM@HOME (<http://boinc.fzk.de>), which explores the free-energy landscape in many parallel dynamical processes, which are coordinated in a single evolutionary algorithm population as outlined in section 2.1. The overall computational work is thus segmented into medium size work-units, which can be processed independently.

3 Results

3.1 Folding of the HIV Accessory Protein 1F4I

The 40 amino acid target 1F4I was folded using the evolutionary algorithm on POEM@HOME. The population was initially seeded with a single extended 'stick' conformation. Figure 2 shows the convergence of the energy as a function of the total number of basin hopping cycles. We find that the best energy converges quickly to a near-optimal value with the total number of basin hopping cycles. As a result of the population diversity criterion, there will always be a finite difference between the average and best energy. This is established by an acceptance threshold of 3 Å RMSD for the inclusion of new structures into the population.

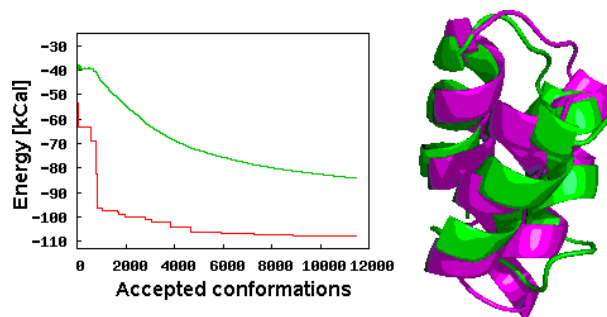


Figure 2. Left: Evolution of average (upper curve) and best (lower curve) energy in the folding process. Right: Overlay of the simulated and experimental structure of 1F4I.

E [kcal]	RMSD [Å]	Secondary Structure
Exp	-	CCHHHHHHHHHHTTCCCHHHHHHHHHHTTTSCSHHHHHHHHHHC
-107.77	2.56	CHHHHHHHHHHHHSCCHHHHHHHHHHHHHCHHHHHHHHHHC
-107.12	8.11	CHHHHHHHHHHHSCSSSSCBTTSCCSHHHHHHHHHSCSBC
-106.30	6.60	CHHHHHHHHHHHHCSSSHHHHHHHHHHHCHHHHHHHHHHC
-103.90	7.95	CCHHHHHHHHHHSCCSSSEEBTTBCSSHHHHHHHHHCCCEEC
-103.66	4.90	CCHHHHHHHHHHCCCHHHHHHSCCBTTTBHHHHHHHHHHHC

Table 1. Top five best energy structure of different topology for folding Target 1F4I.

Another indication of the diversity of the algorithm can be observed in table 1. All of the best energy structures show significant differences in their secondary structures. The best energy-structure found has a RMSD of 2.56Å to the native structure.

4 Conclusions

We have shown, that the mapping of the 'folding problem' onto an optimization problem permits the use of methods that speed the exploration of the free-energy surface. The present study demonstrates, equally importantly, that it is possible to parallelize the search process by splitting the simulation into a large number of independent conformations, rather than by parallelizing the energy evaluation.

Acknowledgments

We thank the Fond der Chemischen Industrie, the BMBF, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-1, WE 1863/10-2 and WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. We also thank the BOINC (<http://boinc.berkeley.edu>) developers and especially our current POEM@HOME Users for their support.

References

1. A. Verma, W. Wenzel, *All-atom protein folding using an improved free-energy force-field*, Biophys. J: Suppl. **S**, 214, 2007.
2. T.Herges, W. Wenzel, *An all-atom forcefield for tertiary structure prediction of helical proteins*, Biophys. J **87**, 3100, 2004.
3. C. B. Anfinsen, *Principles that govern the folding of protein chains*, Science **181**, 223–230, 1973.
4. T.Herges, W.Wenzel, *In Silico Folding of a Three Helix Protein and Characterization of Its Free-Energy Landscape in an All-Atom Force Field*, Phys.Rev.Lett. **94**, 018101, 2005.
5. A.Verma, A.Schug, K.H. Lee, W.Wenzel, *Basin hopping simulations for all-atom protein folding*, J.Chem.Phys. **124**, 044515, 2006.
6. A.Schug, W.Wenzel., *Predictive in-silico all-atom folding of a four helix protein with a free-energy model*, J. Am. Chem. Soc. **126**, 16736–16737, 2004.
7. A.Schug, W.Wenzel., *An evolutionary strategy for all-atom folding of the sixty amino acid bacterial ribosomal proein L20*, Biophys. J. **90**, 4273–4280, 2006.